Short-term effects of sound localization training in virtual reality

Mark A. Steadman[1,2*], **Chungeun Kim**[1], **Jean-Hugues Lestang**[3], **Dan F. M. Goodman**[3], **and Lorenzo Picinali**[1]

[1]Dyson School of Design Engineering, Imperial College London, London, UK
[2]Department of Bioengineering, Imperial College London, London, UK
[3]Department of Electrical and Electronic Engineering, Imperial College London, London, UK
[*]m.steadman@imperial.ac.uk

ABSTRACT

Head-related transfer functions (HRTFs) capture the direction-dependant way that sound interacts the head and torso. In virtual audio systems, which aim to emulate these effects, non-individualized, generic HRTFs are typically used, leading to inaccurate virtual sound localization. Training has the potential to exploit the brain's ability to adapt to these unfamiliar cues. In this study, three virtual sound localization training paradigms were evaluated; one provided simple visual positional confirmation of sound source location, a second introduced game design elements ("gamification") and a final version additionally utilized head-tracking to provide listeners with experience of relative sound source motion ("active listening"). The results demonstrate a significant effect of training after a small number of short (12-minute) training sessions, which is retained across multiple days. Gamification alone had no significant effect on the efficacy of the training, but the inclusion of active listening resulted in a significantly greater improvement in virtual sound localization accuracy. Improvements in polar angle judgement were significantly larger for the trained HRTFs, while improvement in lateral judgements and front-back reversals also generalized to a second set of HRTFs, for which no positional feedback was given. The implications of this on the putative mechanisms of the adaptation process are discussed.

Introduction

Sounds interact with the head and torso in a direction-dependant way. For example, sounds sources to one side will reach the contralateral ear after a longer delay relative to the ipsilateral ear, and with lower intensity. Furthermore, physical interactions with the head and pinnae, the external parts of the ear, introduce spectral peaks and notches, which can be useful for judging whether a sound source is above, below or behind the listener. This direction-dependant filtering is described by Head-Related Transfer Functions (HRTFs). Virtual audio systems are based on the premise that, if the HRTFs for a given listener are known, any monoaural sound can be processed in such a way that, when presented over headphones they are perceived as if they emanate from any position in 3D space[1].

Because of individual differences in the size and shape of the head and pinnae, HRTFs vary from one listener to another. It follows that an ideal virtual audio system would make use of individualized HRTFs. This is problematic for virtual audio systems that are designed for use in consumer or clinical applications, because the equipment required to measure HRTFs is typically bulky and costly. Some work has been done on estimating HRTFs from readily accessible anthropometric information; for example, measurements of the pinnae and head[2,3] or even photographs[4,5]. However, such approaches necessitate the use of simplified morphological models, the limitations of which are unclear. The most accurate estimations of HRTFs typically involve the use of specialized equipment, ranging from rotating listening platforms to spherical loudspeaker arrays and robotic armatures (for a brief overview see[6]) along with miniature, accurate microphones that can be placed inside the ear. For this reason, consumer-oriented systems typically use generic HRTFs measured from a small sample of listeners, or artificial anthropometric models such as the KEMAR head and torso[7].

It is generally thought that the differences between individualized HRTFs and these generic ones have a detrimental effect on the accuracy and realism of virtual sound perception. It has been noted, for example, that listeners are able to localize virtual sounds that have been spatialized using their own HRTFs with a similar accuracy to free field listening (albeit with somewhat poorer elevation judgments and increased front-back confusions[1,8]). These errors are exacerbated by the use of non-individualized HRTFs[9,10]. Furthermore, it has been suggested that non-individualized HRTFs result in an auditory perception with reduced "presence"[11].

It would seem that the efficacy of virtual audio systems utilising generic HRTFs is limited by the perceptual similarity between those HRTFs and the listener's own. Indeed, efforts have been made to "match" listeners to optimal HRTFs from a database using subjective methods[12–14]. Whilst this is a promising, efficient approach, it does not take advantage of the brain's ability to adapt to changes in sensory input. There is increasing evidence that the adult brain is more adaptable than classically thought[15]. For example, it has been demonstrated that this adaptability (or plasticity) can lead to a decrease in localization error over time when a listener's normal cues for sound location are disrupted by physically altering the shape of the ear using molds[16–18]. However, this process typically takes place over the course of days or weeks (for a review see Mendonça *et al.*[19]).

Such timescales are likely to be impractical for a consumer-oriented or clinical applications, where rapid optimization is generally desirable. The possibility of accelerating the process of adapting to "new ears" has therefore received some attention. Encouragingly, several studies have demonstrated that training through positional feedback (for example, indication of virtual sound source location using visual or somatosensory cues) has the potential to achieve adaptation over timescales of the order of a few hours or even minutes[20–24]. Whilst it seems clear that explicit training can result in better outcomes in virtual audio, whether measured by localization accuracy or perceived externalization, improvements over short timescales are typically small and highly variable.

It is possible that such training paradigms could be further optimised. One promising avenue of investigation to that end is the use of "gamification", whereby design elements traditionally used in gaming are employed in a non-gaming context. Not only is gameplay engaging, having the potential to improve attention to a perceptual learning task, but it also leads to the release of reward signals such as dopamine[25], which in turn have been purported to have an enhancing effect on perceptual learning through the promotion of synaptic plasticity in sensory processing nuclei[26]. The efficacy of video games to enhance various aspects of perceptual learning has been explored in the visual domain[27–30] and, more recently, in the auditory domain[31–34]. However, to what extent gamification can accelerate virtual sound localization training relative to a more traditional approach is unknown.

Changes in virtual sound localization performance following training are typically described as "HRTF adaptation"[19, 23, 35]. However, the mechanisms underlying this adaptation are unclear. One possibility is that this adaptation reflects a process of learning a new, internal representation of the unfamiliar HRTFs in parallel to the listener's own[16]. In this case, one would expect any changes in localization performance to be quite specific to the HRTF used during the adaptation or training period. A second possibility is that the process may involve cue reweighting, whereby the listener learns to prioritise cues that remain robust despite perceptual differences between their own HRTFs and the generic set. If this is the case, listeners would be likely to prioritise cues that generalise to several generic HRTF sets. A cue reweighting mechanism has been reported in adult listeners in a sound localization study utilizing unilateral ear plugs[36]. Understanding the mechanism of HRTF adaptation could have implications for virtual audio system design and may be of interest in the field of auditory perceptual learning more generally.

The first aim of this study was to develop and compare several training paradigms that could be used to facilitate and measure adaptation to non-individualized HRTFs. We explored the efficacy of gamification for this purpose and contrasted it with a more traditional psychophysics protocol. We also directly addressed the question of whether "active listening" (moving the head relative to a fixed virtual sound source) is an important factor in adapting to "new ears". Previous studies have compared improvements due to training for listeners using their own, individualized HRTFs with those using non-individualized HRTFs[23, 37]. Here, we aimed to investigate whether any learning effects as a result of training generalized to a second set of non-individualised HRTFs, for which the listeners received no positional feedback. Finally, since this study was, in part, motivated by the requirement to implement effective virtual audio without bulky, expensive specialized equipment, this was achieved using readily available consumer electronics.

Results

In total, 36 participants were recruited for this study. These were divided into four groups, three of which underwent sound localization training; non-gamified ($n = 9$), gamified ($n = 7$), or active listening ($n = 11$). The final group acted as a control and did not undergo training ($n = 9$).

Effects of virtual sound localization training

Virtual sound localization errors were measured before and after localization training using sounds spatialized using non-individualized HRTFs presented over headphones. Participants were required to indicate the perceived direction of the virtual sound sources by orienting towards it while their orientation was tracked using embedded sensors in a smartphone-based, head mounted display. Between testing blocks, participants underwent virtual sound localization training, during which they were provided with visual positional feedback indicating the sound source location after giving a response. There was a total of nine, 12-minute training blocks split over three days. Additional testing blocks were carried out at the beginning and end of each day, and between every training block on the first day in order to capture the dynamics of any very rapid changes in localization accuracy. This section presents the changes that occurred over the entire course of training. The timescale of learning is addressed explicitly in a subsequent section.

A total of 36 participants were randomly assigned to one of four groups; a control group ($n = 9$) that did not complete any training blocks, and three training groups. Each training group utilised a different version of the training software, referred to as non-gamified ($n = 9$), gamified ($n = 7$), and active-gamified ($n = 11$). The distributions of localization errors, as measured by the angle between the target and response orientations, are shown in the top row of Fig. 1. In all groups and testing blocks, the errors tend to be skewed, with the majority of errors having a magnitude of <90°, with some extending to almost 180°(the maximum). For this reason, per-participant median errors are used in subsequent statistical analyses.

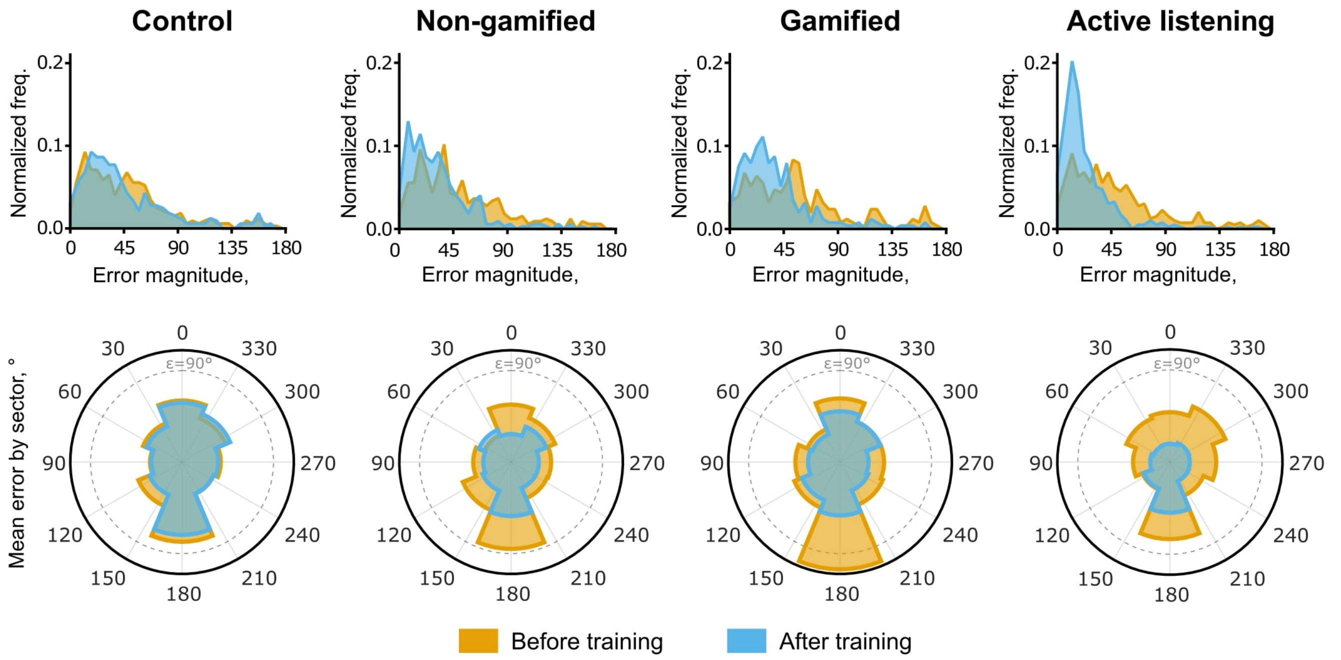

Figure 1. (Top row) distribution of localization errors pooled across all participants within each group before training (orange) and after completing a total of nine, 12-minute training sessions across three days, or following a matching evaluation schedule without training for the control group (blue). (Bottom row) polar histograms of average localization error grouped by target azimuth into eight sectors both before (orange) and after training (blue). The dashed lines indicate a mean localization error of 90°.

The bottom row of Fig. 1 shows the localization error as a function of target azimuth before and after training. In all cases, the largest errors were observed when the virtual sound sources were located directly behind the participants. Although the errors are substantially reduced following training in all groups except the control, the largest errors still tend to occur for targets in this region.

The data shown in Fig. 1 suggest that that the localization training had the effect of reducing overall localization error, since there appear to be more pronounced changes from before to after training for each of the trained groups relative to the control group. To measure the efficacy of localization training generally, data were initially pooled across the trained groups and compared to the control group. The per-participant median errors were used in this analysis. The localization errors for the control ($n = 9$) and pooled trained ($n = 27$) groups, before and after training, are summarized in Fig. 2a. These groups were well matched in terms of initial localization error ($\mu_{control} = 42.6$; $\mu_{trained} = 43.3$). For the control group, the mean error decreased by 6.25°from the initial to the final evaluation. However, a paired-samples t-test indicated that this difference was not statistically significant, $t(16) = 1.02; p = 0.322$. For the pooled trained groups, the mean error decreased by 18.4°. A similar, paired samples t-test revealed this difference to be statistically significant, $t(52) = 6.56; p < 0.001$.

In order to directly examine the effect of training vs no training, a one-way ANCOVA was conducted to compare the final localization errors for the trained and control participant groups, whilst controlling for and differences in initial localization error. This revealed a significant effect of training on final localization error after controlling for initial localization error, $F(1,33) = 13.4, p < 0.001$. This is illustrated in Fig. 2b, which shows that, for a given median error before training, the median error after training is generally lower for the trained group.

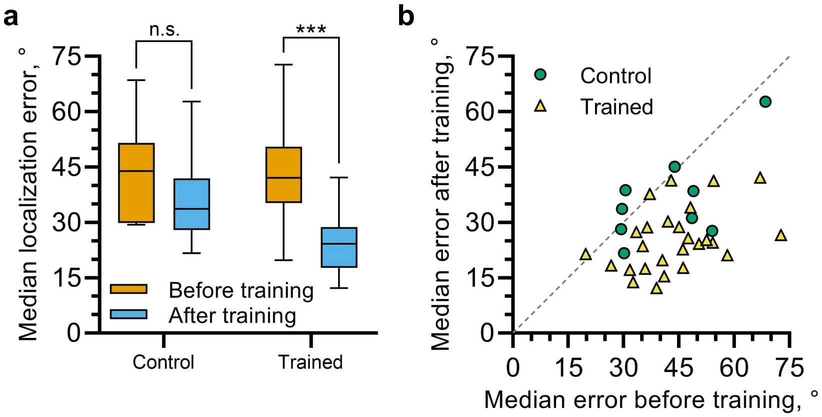

Figure 2. (**a**) distribution of per-participant median localization error for the control (left) and pooled trained (right) participant groups. The errors during the initial (orange) and final evaluation sessions (blue) are shown, corresponding to no training and following nine 12-minute training sessions across three days. (**b**) initial vs final localization errors for each participant in the control (green ◯) and pooled trained (yellow △) groups. The gray, dashed line indicates parity between initial and final localization errors.

Changes in lateral and polar angle judgements

The above results demonstrate an improvement in overall localization accuracy as a result of training with positional feedback. These improvements could be driven by timing or level differences between the ears, which serve as dominant cues for left-right position, as well as spectral cues, such as the position of spectral notches, which serve as a cue for source elevation and facilitate resolution of front-back ambiguity. In order to investigate the relative contributions of these factors to the overall reduction in localization error, we used the auditory-inspired interaural polar coordinate system[8] to define errors in terms of their lateral, polar and front-back components.

One-tailed paired samples t-tests were conducted to compare initial (before training) and final (after training) values of the lateral, polar and front-back confusion measures separately. Data were pooled across the three training groups and are summarized in Fig. 3. In order to separate polar angle errors from front-back confusions, all targets and responses were projected onto the front hemisphere before calculating the polar angle error. All tests revealed a significant reduction in error in the final evaluation compared to the initial evaluation; lateral: $t(26) = 3.75, p < 0.001$; polar: $t(26) = 5.93, p < 0.001$; front-back confusions: $t(26) = 6.54, p < 0.001$. A measure of effect size, Cohen's d, confirmed that largest effect was in the reduction of polar angle error, $d_{\text{lateral}} = 0.87$, $d_{\text{polar}} = 1.47$, $d_{\text{front-back}} = 1.08$.

Effects of gamification and active listening

Participants who underwent localization training were divided into three groups. Each group used a different version of the sound localization software; a basic version ("non-gamified"; similar to the testing procedure except visual positional feedback was provided), a "gamified" version (incorporating traditional game design elements such as point-scoring and level progression) and a modified version of the gamified software, that encouraged "active listening" by allowing the participant to move their head relative to the virtual sound sources during playback. We hypothesized that both the gamified and active listening variations would lead to greater reductions in localization error than the non-gamified version.

The overall localization errors before and after training separated by training type are summarized in Fig. 4. Separate paired t-tests confirmed that each training paradigm resulted in statistically significant reductions in localization error (Fig. 4a; $t_{non-gamified}(8) = 4.94, p < 0.0011$; $t_{gamified}(6) = 6.33, p < 0.0115$; $t_{active}(10) = 6.33, p < 0.001$). In order to directly compare the training types, a one-way ANCOVA was conducted to determine whether there was a statistically significant difference between the final localization errors for the training groups, whilst controlling for any differences in initial localization error (Fig. 4b). This revealed a significant effect of training type, $F(2,23) = 8.43, p = 0.00179$. Tukey *post hoc* tests revealed that the adjusted final localization error was significantly lower for the active listening group compared to both the non-gamified ($\Delta_{error} = 7.15°, t = 2.54, p = 0.0464$) and gamified groups ($\Delta_{error} = 12.6°, t = 3.97, p = 0.00173$). There was no significant difference between the non-gamified and gamified groups.

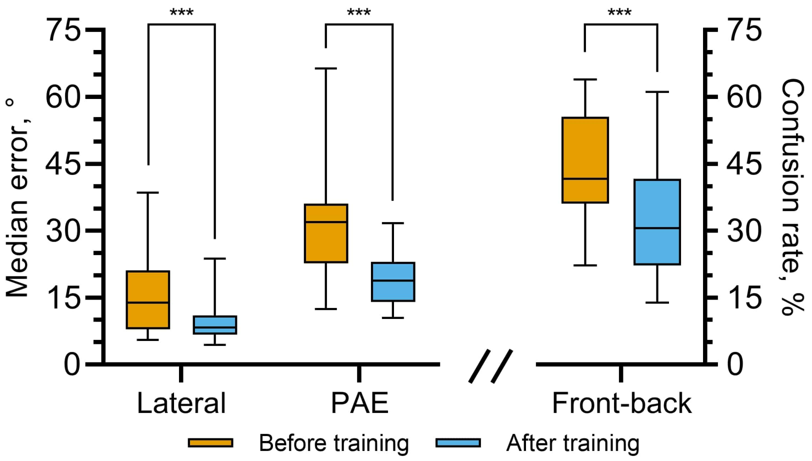

Figure 3. Distribution of per-participant median errors separated out by error type before training (orange) and after nine, 12-minute training sessions over three days (blue). Data are from all participants who underwent training, pooled across the training paradigms. Lateral and polar errors refer to the interaural-polar coordinate system (see Methods) and are presented in degrees (left y-axis). Polar angles are calculated after all targets and responses were projected to the front hemisphere. Front-back confusions are presented as a percentage (right y-axis).

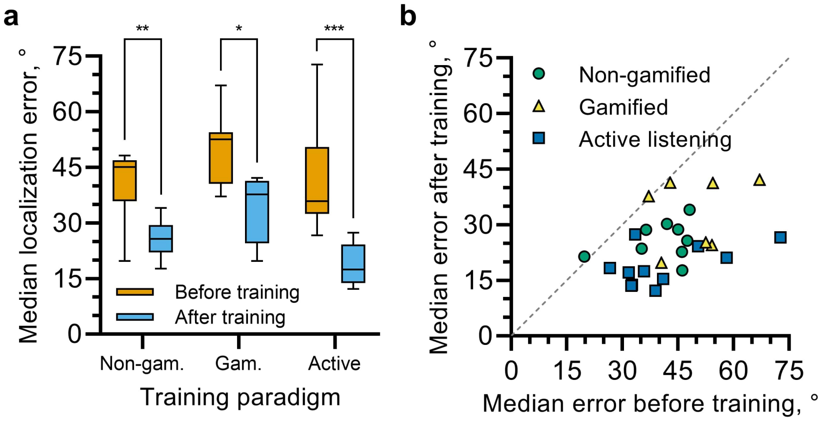

Figure 4. (**a**) distribution of per-participant median localization error for each of the participant groups undergoing localization training (Non-gam = Non-gamified, Gam. = Gamified, Active = Active listening). The errors during the initial (orange) and final evaluation sessions (blue) are shown, corresponding to no training and following nine 12-minute training sessions across three days. (**b**) initial vs final localization errors for each participant in trained groups. The gray, dashed line indicates parity between initial and final localization errors.

Timescale of learning

To investigate the timescale of the adaptation process, with a particular focus on any short-term learning effects, testing blocks were carried out between each of the three, 12-minute training block on the first day, and following each block of three training blocks on days two and three. The localization errors at each of these time points are shown in Fig. 5a. Again, these data are pooled across the three trained groups. The change in total localization error as a function of the number of training sessions carried out was well described by an exponential decay of the form $aexp(-bx)+c$, where x corresponds to the number of completed training sessions and a, b and c were parameters determined using the MATLAB function `fit`. In this form, the parameters a and c define to the initial and final localization error and b relates to the rate of learning. The optimal parameters were found to be $a = 18.15$, $b = 0.364$ and $c = 24.4$. This corresponds to an initial error of $a + c(42.5°)$, a final error of 24.4°, and suggests that errors will reduce by half following 1.90 training sessions (equal to $\ln(2)/b$), equating to approximately 23 minutes.

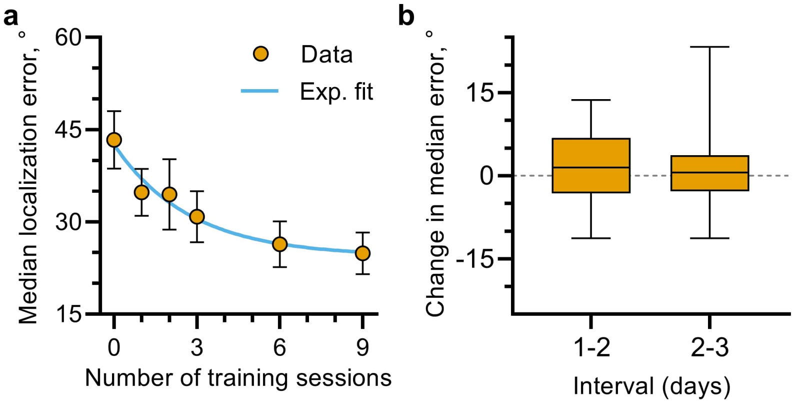

Figure 5. (**a**) Change in total localization error as a function of the number of 12-minute training sessions. The data are pooled across all participants undergoing training and reflect the per-participant median spherical angle errors. The blue line shows a fitted decaying exponential function. (**b**) Change in per-participant median error between sessions on consecutive days. No training took place during these intervals.

Testing was additionally carried out at the beginning of days two and three, following a gap of up to two days following the previous training session. This facilitated the examination of any consolidation or latent learning. These data are shown in Fig. 5b, which illustrates the change in localization error from the last testing block on days one and two and the initial testing block on days two and three respectively. Separate t-tests were carried out on the change in localization error from days one to two and days two to three. The mean change in error from the end of one day to the start of the next was marginally greater than zero for both intervals, $\mu_{1-2} = 1.67°, \mu_{2-3} = 2.52°$. However, the separate t-tests suggested that these differences were not significant, $t_{1-2}(25) = 1.21, p_{1-2} = 0.236$; $t_{2-3}(25) = 1.62, p_{2-3} = 0.117$, indicating that localization accuracy remained fairly consistent between sessions.

Changes in response bias

A large proportion of the improvements in localization appear to occur very rapidly during the first day of training. One possibility is that a rapid shift in response bias might account for this change. Fig. 6 shows the locations of responses given by four example participants showing clear response biases during the initial testing block (top row). These participants also illustrate the high degree of variability in response biases. For example, participant I shows a bias towards the front, giving responses in the rear hemisphere only rarely, while participant II shows the opposite pattern. The responses of participant III lay predominantly close to the sagittal plane, while those of participant IV were well distributed across target azimuth but constrained to low elevation angles. The distribution of responses is much more even following a single, 12-minute training block (middle row). Biases are no longer apparent by the end of training (bottom row).

Generalizability to an untrained HRTF

The previous sections have shown that each of the virtual sound localization training paradigms led to a reduction in localization errors for sound spatialized using a set of non-individualized HRTFs. To explore whether this learning effect was specific to only this HRTF set, target sounds were also spatialized using a second set of HRTFs within each testing block. As previously indicated in Fig. 2a, the mean localization error for sounds spatialized using the trained HRTFs decreased by 18.4°following training. Interestingly, the mean localization error for sounds spatialized using the non-trained HRTFs also decreased, $\Delta\mu = 15.9°$.

The changes in localization error following training for both the trained and non-trained HRTFs are summarized in Fig. 7a. To test if the changes differed for the two HRTFs, separate, paired t-tests were carried out on the each aspect of localization error. No significant difference between the HRTFs was found in overall change in spherical angle error, $t_{\text{spherical}}(26) = 1.00, p = 0.325$. This was also reflected when looking at only the lateral and front-back confusion aspects of this error individually; $t_{\text{lateral}}(26) = 0.277, p = 784$; $t_{\text{front-back}}(26) = 0.150, p = 0.882$. The polar angle error did reveal a

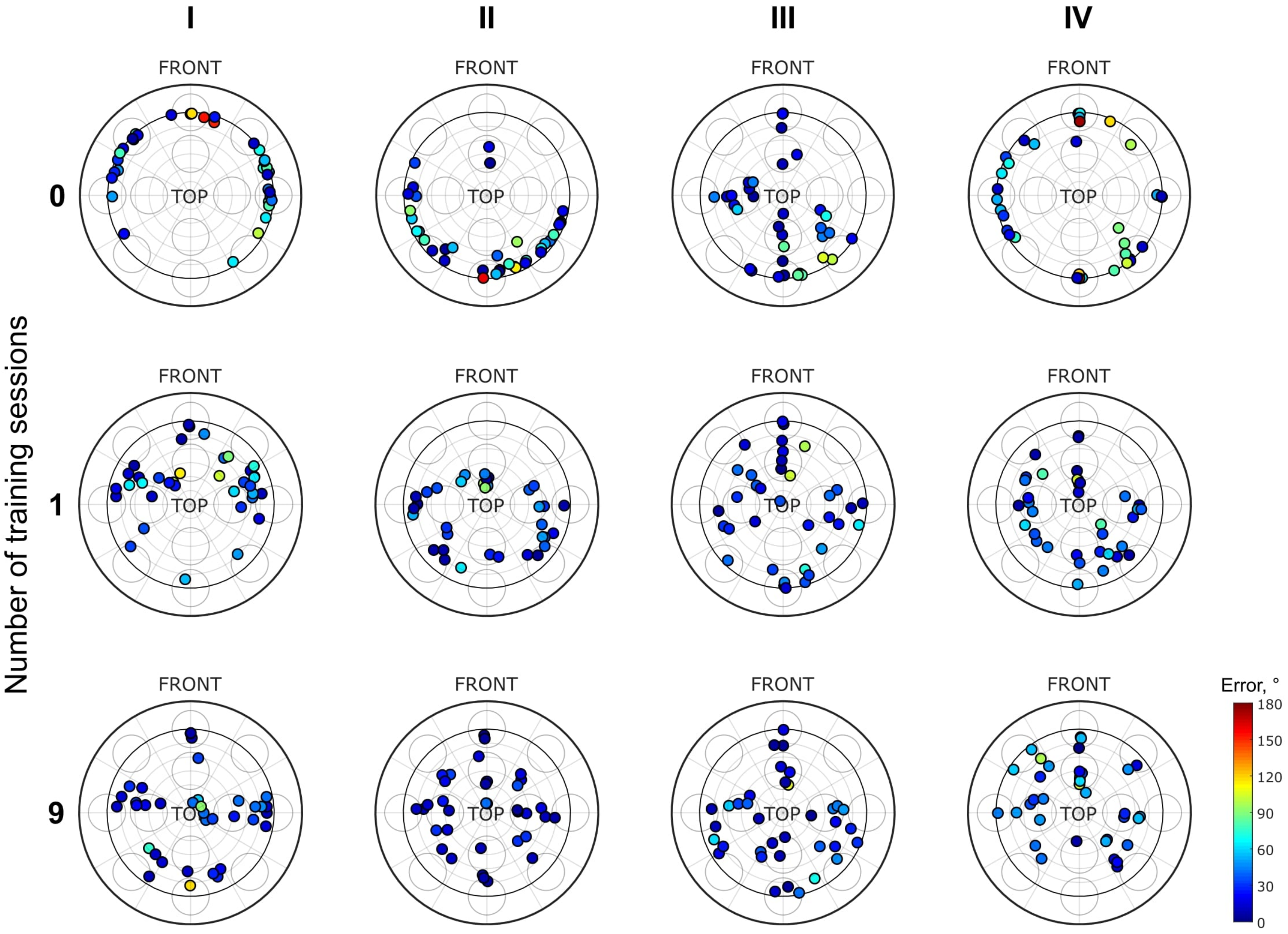

Figure 6. Changes in response biases from before training (top row) to following a single, 12-minute training session (second row) and following the final training session (bottom row). Each dot indicates the orientation of a response with azimuth indicated by the polar angle from straight up on the axes. Elevation is indicated by the distance from the centre, with the origin corresponding to straight up and the inner dark ring indicating 0°elevation. The magnitude of the spherical angle error for each response is indicated by the colour of each dot (see colourbar).

significant difference between the HRTFs $t_{\text{polar}}(26) = 3.01, p = 0.0057$. However, as previously noted this difference was too small to contribute to a significant difference in localization improvements overall.

In order to account for any differences in pre-training localization accuracy for the two HRTF sets for each participant, these data were analyzed using separate ANCOVAs for each aspect of the localization error. These ANCOVAs treated the HRTFs (trained or non-trained) as a fixed factor with the median post-training error as the dependant variable and median pre-training error as a covariate. This analysis corroborated those described above, whereby the spherical angle, lateral angle and front-back confusion analyses indicated no significant differences between the two HRTFs ($F_{\text{spherical}}(1,51) = 0.691, p < 0.410$, $F_{\text{lateral}}(1,51) = 0.898, p < 0.348$, $F_{\text{front-back}}(1,51) = 0.031, p < 0.861$). Analysis of the polar angle error, however, did reveal a significant difference between the post-training error while accounting for difference in pre-training performance ($F_{\text{polar}}(1,51) = 0.17.135, p < 0.001$).

In summary, these data show that the training effect tended to generalize to a second set of non-individualized HRTFs, for which no positional feedback was given, although HRTF-specific improvements can be observed in judgements of polar angle where front-back confusions are ignored.

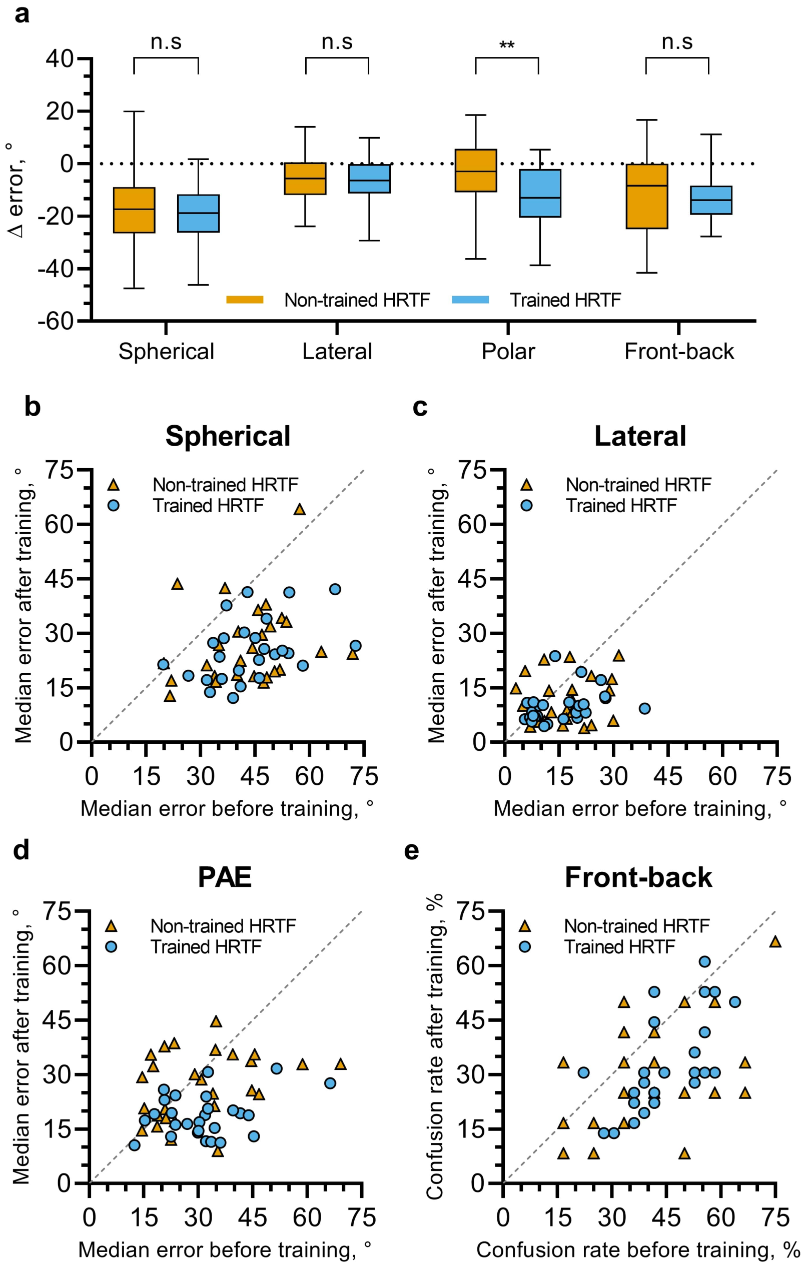

Figure 7. Difference in overall change in localization error following nine, 12-minute training sessions for sounds spatialized using the same HRTFs used in the training (trained HRTF) and a second set of HRTFs, not used in the training (non-trained HRTF). **a** shows the overall changes in total, spherical angle error as well as the lateral, polar and front-back aspects of this error. Data are pooled across training types. **b-e** Show the errors before and after training for each particiaptn individually.

Discussion

It is now well established that the adult human brain can adapt to modified or unfamiliar cues for sound direction. This has been demonstrated in studies using ear plugs[36,38] as well as molds inserted in the pinnae[16,17]. Some recent studies have also demonstrated a learning effect using visual positional feedback delivered via a head-mounted display. Zahorik *et al.*[20] used a very similar training paradigm to those described here and found that the majority of the listeners demonstrated improvements in front-back judgements following two, 30 minute training sessions. However, they reported no significant changes in lateral or elevation judgements. Conversely, Majdak *et al.*[21] reported more general improvements in virtual sound localization precision following training with visual positional feedback, this time using individualized HRTFs. Similarly, Parseihian and Katz[23] noted improvements in lateral and elevation judgements. Interestingly, this was only the case for the group of listeners using "poorly matched" HRTFs (but not in those using individualized, or "well matched" HRTFs).

The task of quantitatively comparing across studies is non-trivial for several reasons. Firstly, a ubiquitous finding seems to be the large degree of variation across listeners both in localization ability before training and in the effectiveness of the training, even when initial ability is taken into account. This makes it highly likely that findings will be inconsistent, especially where small sample numbers are used. Secondly, differences in pointing method and the distribution and and nature of the target stimuli are all likely to have a significant effect on various measures of localization accuracy. Nonetheless, the localization errors before training in this study seem to be consistent with published results where direct comparison is tractable. For example, the average lateral error across all participants before training was 15.1°. This is very close (within ±1°) to those reported by Zahorik *et al.*[20], and Majdak *et al.*,[39], which both made use of head-mounted displays and virtual environments. The mean front-back confusion rate was initially very high at 44.6%. However, similar results have been reported in other studies using non-individualized HRTFs[9,20], and this is also in the context of high variability between individuals. In summary, these comparisons at least suggest that the participants recruited here were unremarkable in terms of localization ability.

This study aimed to explore the potential of "gamification" to enhance the effects of training on adaptation to non-individualized HRTFs. It has previously been shown that videogame-like paradigms can be used to effect improvements in virtual sound localization[31]. Here, we compared a standard, adaptive psychoacoustic training paradigm to "gamified" version of the same task, which incorporated performance-related feedback by, amongst other things, awarding points for positive actions ("hits") and decreasing player "health" for negative actions ("misses"). However, the introduction of these elements alone was insufficient to effect a significantly better outcome than the more traditional version.

It has been previously shown that game-play can have an enhancing effect on perceptual learning, even where the video game is not directly related to the learning task, or indeed involving the same modality[34,40]. However, it may be that our "gamified" version was not sufficiently distinct from the standard task, since both versions provided performance-related feedback in form of "hits" and "misses" and, anecdotally, participants reported being quite motivated to improve over time. Since perceptual learning enhancement appears to be effected through videogame play regardless of relation to the learning task, it is possible that an alternative approach might be to incorporate existing games, or at least those designed primarily for fun, into the localization training regimen, rather than designing the game around the localization task.

Active listening, on the other hand, does seem to lead to a more marked improvement. Here, active listening refers to the ability of the listener to experience a stationary virtual sound source while being free to move their head. This necessitates the use of head tracking, whether that is through a dedicated head-tracking system, or simply using the positional data from embedded sensors in a smartphone (as used here). The results from this study suggest that the experience of continuous, relative motion may have an enhancing effect on virtual sound localization training. However, it should be noted that the active listening group were presented with the stimuli repeatedly during a trial so it could be argued that this may simply be an effect of greater exposure.

In comparison to many previous studies on adapting to altered cues for sound localization, this study focussed on changes over a very short timescale. Typically, such adaptation has been measured over the course of hours, days and weeks (for a review see Mendonça *et al.*[19]). In agreement with the few studies that have also focussed on short-term changes, we show that significant changes occur after a total training duration of the order of an hour[20,22–24]. This is an encouraging result if future systems using generic, non-individualized HRTFs are to find broader application, since it seems likely only a minority of use cases will justify longer adaptation periods.

The change in localization error was well characterised by a decaying exponential with the greatest changes occurring very rapidly. Such rapid changes are often attributed to procedural learning, the specific mechanisms of which are poorly defined, but typically encompass any aspects of learning other than those specific to the stimulus or task (e.g. familiarisation with the testing method or response demands[41]). However, there is some evidence that perceptual learning can also have a very rapid onset and changes during this initial period reflect a combination of both perceptual and procedural learning[42]. Attempts were made to mitigate the effects of procedural learning during the early stages of training by including an acclimatisation "tutorial" phase in which participants were familiarised with the interface, virtual environment and pointing method. Furthermore, learning effects were not constrained only to the very initial stages during which one would expect effects of familiarisation to be most profound, suggesting that the improvements demonstrated here were unlikely to be purely a result of procedural learning.

Since localization testing took place at the beginning and end of each training day, it was also possible to test for any consolidation, whereby the effects of training remain robust over extended time periods, or between-session learning effects, whereby the effect of training is exacerbated following a rest period. Here, a consolidation effect was observed whereby improvements were maintained between sessions. But we observed no between-session latent learning as has been observed in a previous study on auditory perceptual learning, albeit on a simpler task[43]. This study concluded that very short training sessions (∼8 minutes) can optimise latent learning and can therefore be the most efficient regimen. It is possible that shorter training sessions spread over more days might, indeed, be optimal and could be appropriate in some use cases.

A key difference between the study design used here and previous, similar studies on adaptation to non-individualized HRTFs is that listeners were also tested using a second set of non-individualized HRTFs for which no positional feedback was given. This facilitated measurement of the extent to which the learning effect was specific to the trained HRTFs. Previous studies have shown that localization training paradigms can lead to learning effects that generalize to novel stimulus-position pairings[24], but to our knowledge generalization across HRTFs has not been investigated directly. This has implications for understanding the putative mechanism involved in adapting to non-individualized HRTFs.

One reason one might expect the observed generalization is that the learning effects are primarily due to procedural learning, but as discussed above this seems unlikely to underlie all of the improvements observed. This effect might also be observed if the HRTF sets used in this study happened to be perceptually similar. The method used to select them from the HRTF database would certainly not guarantee perceptual distinctiveness; the subset they were taken from was selected on a somewhat utilitarian basis whereby they tended to produce the most subjectively realistic, spatialized percepts for the greatest number of listeners[14]. It could be argued that this method would tend to produce subsets that epitomize stereotypical features and minimize idiosyncratic variation. To discriminate between changes due to HRTF-specific adaptation and other factors, it would be useful to select HRTFs that are maximally perceptually distinct. This is non-trivial, especially where the cues used for discriminating the specific target orientations used are complex and poorly defined. However, this could be achieved using clustering analyses where HRTFs are specified using dimensions known to be perceptually relevant cues for direction, such as that proposed by So et al.[44].

An alternative possibility, which could account for the observation that the adaptation appears to generalize to more than one HRTF is that the process involves a re-weighting of acoustic cues for sound source location. In this scenario, listeners would learn to rely less on features specific to their own HRTFs and more on features that are shared with the non-individualized ones. An illustrative example might be that listeners begin to rely on interaural level differences more than time differences for lateral angle judgements, if these are indeed perceptually more robust. Such a mechanism relies on redundancy in auditory-spatial cues and would explain the observation that HRTF adaptation does not have a detrimental effect on localization with listeners' own HRTFs. This has also been put forward as a process underlying auditory perceptual learning in other contexts[36,45]. Such a mechanism also seems to be supported by the rapid change in response bias. It could be that, during the initial testing block, participants did not receive the cues they would normally rely on for the most accurate localization leading to diverse "guessing" strategies. However, upon receiving some positional feedback in a single traing block, they were able to rapidly switch listening modes to place greater emphasis on cues that generalized between their own and the unfamiliar, non-individualized HRTFs. Indeed, such a rapid switch of listening mode can be observed in listeners switching between complex listening environments[46].

One of the aims of this project was to demonstrate that this adaptation could be achieved using readily available, consumer electronics. Previous studies on adapting to non-individualized cues for virtual sound location have utilised specialist facilities and equipment typically involving dedicated head-tracking equipment[23,39] or *ad hoc* visual displays[35]. Whilst dedicated platforms may have the potential for greater experimental control and accuracy, more pragmatic solutions will be required if non-individualized virtual audio is to find more widespread application. Here we demonstrate that the accuracy of localization using consumer equipment is comparable to previous studies and it is possible to effect and measure adaptation to non-individualized HRTFs in the same way.

Since developing the platform used in this study, several systems have become freely available that implement HRTF-based virtual sound spatialization entirely on the mobile device (e.g.[47]), which make it relatively straightforward to implement a similar platform using only a smartphone and a headset. This simplicity raises the possibility that similar virtual sound systems could have broader application than previously thought. For example, it is possible to imagine research, or even clinical, applications whereby the effectiveness of hearing prostheses could be tested in controlled, virtual environments without the need to install bulky, expensive loudspeaker arrays. Of course, the effectiveness of such applications is only assured if it is possible to assume that the reduction in localization error observed following brief periods of training corresponds to an adaptation mechanism whereby the cues and strategies used are similar to those used in real-world listening.

It follows that one limitation of our approach is the use of localization accuracy as a proxy measure for "HRTF adaptation". If the goal is to use the brain's ability to adapt to broaden the application of virtual sound systems either in consumer products, auditory research or clinical applications (such as audiometry in simulated real-world environments), then the quantity we are actually interested in is perceptual "realism". This is ill-defined and difficult to measure reliably[48], which is presumably why localization accuracy is commonly used. The extent to which this proxy measure is associated with perceptual realism, however, is not clear.

In summary, we have demonstrated a virtual sound system that uses only readily available consumer electronics. We implemented training programmes that led to significant decreases in virtual sound localization error following fewer than than two hours of training (split over three sessions). The majority of improvements in sound localization accuracy generalized to a second, non-individualized HRTF set, which was not used in training. Only a small proportion of improvements in polar angle judgements were HRTF-specific. This suggests that the observed improvements in localization may not wholly reflect a process whereby the listener learns a parallel sound-spatial map, and rather suggests that the process may involve switching to a different localization strategy by cue-reweighting, for example. This implies that listeners showing HRTF adaptation as measured by localization accuracy may not be using a naturalistic listening mode, which has implications for possible future applications of virtual audio technology.

Methods

Participants

All procedures were reviewed and approved by the Imperial College Research Ethics Committee. A total of 36 participants, aged between 18 and 38, were recruited for this study. Participants were randomly assigned to one of four groups. Three of the groups underwent sound localization testing and training in a virtual environment, each utilizing a different training paradigm. The fourth group acted as a control and remained in a quiet room between testing sessions. Before entering the study, participants were asked to complete a questionnaire, which revealed no reported cognitive or auditory deficits. All data were anonymized.

Stimuli

A set of 19 acoustically complex stimuli were developed. The sounds were designed such that they were naturalistic, consistent with the virtual environment and contained rich cues for localization. The stimuli comprised a combination of pink (1/f) noise, a segment of speech produced by a male talker and a 1 kHz tone. A schematic of the stimulus is shown in Fig. 8a, and an example is provided (Supplementary Audio). An initial 200 ms noise burst was followed by a one second fragment of continuous Italian speech with low level pink noise, another 200 ms noise burst and, finally, a 200 ms, 1 kHz tone. Each segment was ramped on and off using a 10 ms raised-cosine ramp. The relative levels were set such that the stimulus resembled a short radio communication. From this set, a single stimulus was used only during testing, whilst the other 18 stimuli were selected from randomly during training.

Sounds were spatialized using HRTFs from the IRCAM Listen database[49]. Two HRTFs were randomly selected from a subset of this database, which was determined in a previous study to contain the seven HRTFs that produced the best subjective spatialization for the most listeners[14]. These corresponded to participant numbers IRC0008 and IRC0013 database. The former was used to spatialize sound during training and testing. The latter was used only during testing. All stimuli were generated and stored in 44.1 kHz, 16-bit format.

Spherical coordinate systems

Throughout this manuscript, two spherical coordinate systems are used. The first is a single-pole system whereby sound source coordinates are specified using two coordinates; azimuth, θ and elevation, ϕ. In this coordinate system, the azimuth refers to a rotation around a vertical axis passing through the centre of the head with 0° being directly in front and 90° being directly to the left. Elevation is the angle from the horizontal plane with 0° being horizontal and 90° being directly upwards. This coordinate system is intuitive and mathematically convenient and is therefore used in the subsequent description of target orientation as well as in the experiment software.

A second coordinate system is used in the data analysis, referred to variously in the literature as "head-related"[8], "double-pole"[50] or "interaural-polar"[51]. This coordinate system was inspired by binaural hearing and also involves two coordinates; a lateral coordinate that defines the angle between an "interaural axis", which follows the line linking the listeners ears, and the sound source, and a polar coordinate that defines the angle formed between a perpendicular line linking the sound source and the interaural axis and the horizontal plane. In binaural listening, the dominant cues for the lateral coordinate are considered to be the interaural time and level differences and the dominant cues for the polar coordinate are considered to be spectral since it defines a point on a circular cross section of the "cone of confusion"[52]. See reference[50] for a diagram illustrating these coordinate systems. In this manuscript, we distinguish the coordinate systems using the nomenclature of "azimuth" and "elevation" for the single-pole system, and "lateral" and "polar" for the double-pole, interaural system.

Experiment design and procedure

For each participant, the experiment comprised three sessions, each completed on a different day and separated by no more than two days. Each session incorporated sound localization testing and training blocks in a virtual environment. During a session, participants sat on a freely rotating chair in the centre of a dark, quiet room. The virtual environment was presented using a head-mounted display, and auditory stimuli were presented over headphones. Participants interacted with the experiment software using a gamepad. For details of the equipment, see **Equipment and the virtual sound localization environment**. During testing and training blocks, participants initiated trials in their own time by orienting towards a button within the virtual scene and activating it using the gamepad. Doing so initiated playback of a randomly selected complex auditory stimulus (see **Stimuli**).

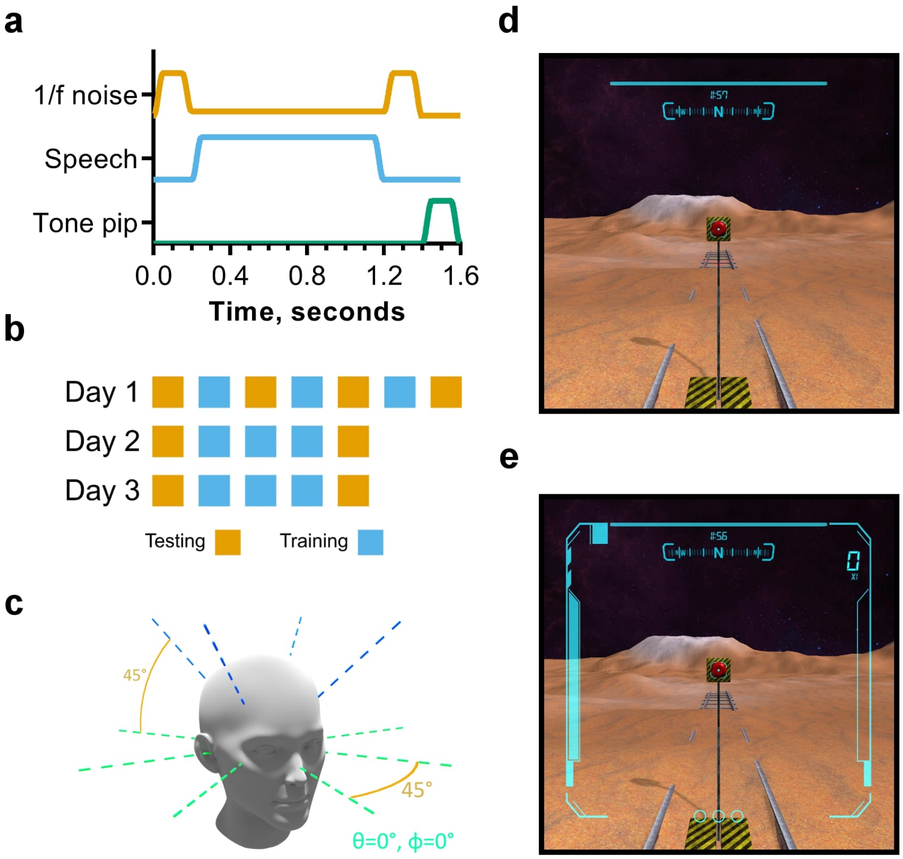

Figure 8. **a**) Schematic representation of the complex stimulus used during training and testing. **b**) schematic of the experimental design indicating the block structure of each session. Testing blocks are in orange and training blocks in blue. Each session was carried out on a different day. **c** Diagram of the centroids of target orientation "regions" used in testing blocks. Target sounds deviated from these centroids by up to 20°. **d,e**) Screenshots of the virtual participants view in the virtual reality application. The marked features correspond to a) timer, b) cardinal direction indicator, c) current score, d) player health indicator, e) animated "charge" indicator visual effect, f) consecutive hit counter. **d** shows the HUD used in the non-gamified version and **e** shows the HUD for the gamified and active-listening versions.

Testing and training blocks were differentiated by whether positional feedback, in the form of a visual sound source positional indicator, was given. Participants underwent an initial testing block at the start of each session, followed by three 12-minute training blocks, followed by a final testing block. Participants were encouraged to take a 5-10 minute break between blocks, during which they remained in the quiet room. In order to capture any very rapid learning effects, additional testing blocks were carried out between each of the training blocks on the first day. This design is represented schematically in Fig. 8b. The control group followed the same process, but remained idle during the periods in which the other groups underwent training.

In training blocks, stimuli were spatialized using randomly generated source locations uniformly distributed over the upper hemisphere by setting $\theta = 2\pi u$ and $\phi = sin^{-1}v$, where θ and ϕ are the azimuth and elevation angles respectively and u and v are random variates uniformly distributed on the interval $[0,1]$. The participants were instructed to remain oriented in the same direction throughout the 1.6 seconds of stimulus playback. If the orientation of the head-mounted display deviated by more than 2.5° in any direction during this time, the trial was cancelled and the training did not continue until the participant initiated a new trial (with a new source location). Following stimulus playback, participants were instructed to orient towards the perceived direction of the sound source and indicate their response using a button on the gamepad.

Visual positional feedback was then provided by introducing a spherical object in the virtual scene at the target sound source location. If the response was within the bounds of the object, the response was indicated to the participant as a "hit", otherwise it was a "miss". The way this was visualized depended on the training paradigm used and is detailed in the following section. The size of the target object varied adaptively. The initial target size was set such that responses were indicated as a "hit" if there was less than 25° deviation from the target in any direction. After achieving three consecutive hits, the target size decreased by 10%. After five misses at a given target size, the target size reverted to the previous one until reaching the initial size. The radius of the target object was therefore given by $r = 0.9^{L-1} d sin\theta$, where L is the current difficulty level, d is the target distance and θ is the allowed angle of deviation for a "hit". This mechanism of "hits" and "misses" was used primarily to provide some performance-related feedback - in all subsequent analyses only the difference between the response and target orientation was used as the variable of interest. All participant groups undergoing training were instructed to initiate and complete trials continuously for 12 minutes per training block.

Testing blocks were carried out using the same virtual environment and process, except positional feedback was not provided following participant responses. Further, testing blocks comprised a fixed number of trials rather than a fixed time limit. In order to ensure consistency across participants, target stimuli were positioned systematically. Eight orientations were defined at 0° elevation, with azimuths were separated by 45° increments beginning at 0°. A further four orientations were defined at 45° elevation with azimuths equal to 45°, 135°, 225° and 315°. These orientations are visualized in Fig. 8c. In order to avoid direct stimulus to response mapping, targets randomly deviated from these orientations by up to 20°. In one testing block, four targets were presented for each target orientation with a different, random deviation each time. Three of these were spatialized using the same HRTFs used in the training blocks. In order to test if learning effects transferred to more than one set of HRTFs, the fourth target was spatialized using a second HRTF set for which the participants had received no positional feedback.

Training paradigms

Three versions of the training software were developed. Each utilised an identical virtual environment. The first, basic version, referred to as "non-gamified", presented participants with a simple head-up display (HUD) incorporating a compass to show absolute orientation as a cardinal direction and a timer, showing the time remaining (Fig. 8d). In this version, visual positional feedback was implemented using plain, spherical objects of uniform colour (green for a "hit" and red for a "miss"). If the target was not in the visual field of the participant when they gave their response, an arrow on the HUD indicated the direction to the target following the shortest path. A second "gamified" version was also developed, which incorporated several videogame design elements. These comprised a score indicator, a consecutive hit counter and player "hit points" (Fig. 8e). Points were rewarded for target "hits". When the target size decreased following the adaptive procedure described above, this was indicated as a level progression to the participant using a sound and text on the HUD (i.e. "LEVEL 2"). In this version, the positional feedback was provided by an animated spherical "robot" in the visual scene, which also fired a laser at the player if a response was a "miss", leading to a decrease in player "health". When the health ran out, this corresponded to an increase in target size, as per the adaptive procedure. A "game over" overlay was presented to the participant and the "hit points" were reset before immediately continuing with training.

In both of these training paradigms, the participant/player was required to keep their head oriented in the same direction during target stimulus playback (within 2.5° deviation) otherwise the trial was cancelled. However, there is some evidence to suggest that "active listening" - the ability of the listener to move their head relative to a sound source - plays an important role in adapting to modified cues for sound localization[19]. A third version of the training software was implemented that incorporated this. There was no requirement for participants to remain oriented in the same direction during stimulus playback and stimulus playback was looped until listeners gave their response. This "active listening" version was in all other ways identical to the gamified version described above. All participants used exactly the same version of the testing software. A video of these training paradigms as well as the testing paradigm is available online (Supplementary Video). Please note that due to the screen capture software and file size limitations, the audio in the video is not representative of the audio generated by the spatialization software.

Equipment and the virtual sound localization environment

The virtual environment was rendered stereoscopically on a smartphone-based (iPhone 6) head-mounted display. Participants interacted with the phone using a handheld controller connected via Bluetooth (Mad Catz C.T.R.L.i). Head-tracking data was transmitted via wireless Ethernet connection to a PC that handled spatial audio rendering. Sound playback and real-time binaural spatialization were implemented using the LIMSI Spatialization Engine[53] - a real-time binaural audio spatialization platform based on Cycling74's Max/MSP. Binaural audio was presented via a Focusrite Scarlett 2i2 USB audio interface using Sony MDR 7506 closed-back headphones. A virtual, moon-like environment was designed to be acoustically neutral to minimize the potential mismatch between the anechoic stimuli and the perceived acoustic properties of the virtual space. The scene was also populated with some landmarks as it has been shown that a lack of visual frame of reference is detrimental to sound localization accuracy[54]. Screenshots of part of the environment are shown in Fig. 8d and Fig. 8e, which show the non-gamified and gamified HUD respectively. The virtual environment can be seen in more detail in the Supplementary Video.

Acknowledgements

This work was supported by the 3D Tune-In project[55], part of the Horizon 2020 research and innovation programme under grant agreement No. 644051. The authors would like to thank Brian F.G. Katz and his team at LIMSI-CNRS for the use of the LIMSI Spatialization Engine.

Author contributions statement

M.A.S, L.P & D.F.M.G developed the study concept and experiment design. M.A.S. designed and build the experimental setup including instrumentation and the virtual environment software. Data was collected by M.A.S., J.H.L & C.H. and analyzed by M.A.S. with input from C.K. Analysis scripts were produced by M.A.S. with input from C.K. M.A.S. produced the manuscript and figures with input from L.P. and D.F.M.G.

Data availability

The Unity project used to build the software for the iPhone 6 and the full dataset, including Matlab scripts to read it, are publicly available on Zenodo[56]. Please note that the authors did not have permission to redistribute the LIMSI Spatialization Engine, which is required to reproduce the stimuli. Please contact the authors for details.

Competing interests

The authors declare no competing interests.

References

1. Wightman, F. L. & Kistler, D. Headphone simulation of free-field listening. ii: Psychophysical. *J. Acoust. Soc. Am* **85**, 868–878 (1989).
2. Kahana, Y., Nelson, P. A., Petyt, M. & Choi, S. Numerical modelling of the transfer functions of a dummy-head and of the external ear. In *Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction* (Audio Engineering Society, 1999).
3. Katz, B. F. Boundary element method calculation of individual head-related transfer function. i. rigid model calculation. *The J. Acoust. Soc. Am.* **110**, 2440–2448 (2001).
4. Dellepiane, M., Pietroni, N., Tsingos, N., Asselot, M. & Scopigno, R. Reconstructing head models from photographs for individualized 3d-audio processing. In *Computer Graphics Forum*, vol. 27, 1719–1727 (Wiley Online Library, 2008).
5. Torres-Gallegos, E. A., Orduna-Bustamante, F. & Arámbula-Cosío, F. Personalization of head-related transfer functions (hrtf) based on automatic photo-anthropometry and inference from a database. *Appl. Acoust.* **97**, 84–95 (2015).
6. Katz, B. F. & Begault, D. R. Round robin comparison of hrtf measurement systems: preliminary results. In *Intl. Cong. on Acoustics 19*, 1–6 (2006).
7. Burkhard, M. & Sachs, R. Anthropometric manikin for acoustic research. *The J. Acoust. Soc. Am.* **58**, 214–222 (1975).
8. Morimoto, M. & Aokata, H. Localization cues of sound sources in the upper hemisphere. *J. Acoust. Soc. Jpn. (E)* **5**, 165–173 (1984).
9. Wenzel, E. M., Arruda, M., Kistler, D. J. & Wightman, F. L. Localization using nonindividualized head-related transfer functions. *The J. Acoust. Soc. Am.* **94**, 111–123 (1993).
10. Begault, D. R. & Wenzel, E. M. Headphone localization of speech. *Hum. Factors* **35**, 361–376 (1993).
11. Väljamäe, A., Larsson, P., Västfjäll, D. & Kleiner, M. Individualized head-related transfer functions, and illusory ego-motion in virtual environments. *Self-motion Presence Percept. Optim. a Multisensory Virtual Real. Environ.* 39 (2005).
12. Seeber, B. U. & Fastl, H. Subjective selection of non-individual head-related transfer functions (Georgia Institute of Technology, 2003).
13. Iwaya, Y. Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears. *Acoust. science technology* **27**, 340–343 (2006).
14. Katz, B. F. & Parseihian, G. Perceptually based head-related transfer function database optimization. *The J. Acoust. Soc. Am.* **131**, EL99–EL105 (2012).
15. Fuchs, E. & Flügge, G. Adult neuroplasticity: more than 40 years of research. *Neural plasticity* **2014** (2014).

16. Hofman, P. M., Van Riswick, J. G. & Van Opstal, A. J. Relearning sound localization with new ears. *Nat. neuroscience* **1**, 417 (1998).

17. Van Wanrooij, M. M. & Van Opstal, A. J. Relearning sound localization with a new ear. *J. Neurosci.* **25**, 5413–5424 (2005).

18. Carlile, S., Balachandar, K. & Kelly, H. Accommodating to new ears: the effects of sensory and sensory-motor feedback. *The J. Acoust. Soc. Am.* **135**, 2002–2011 (2014).

19. Mendonça, C. A review on auditory space adaptations to altered head-related cues. *Front. neuroscience* **8**, 219 (2014).

20. Zahorik, P., Bangayan, P., Sundareswaran, V., Wang, K. & Tam, C. Perceptual recalibration in human sound localization: Learning to remediate front-back reversals. *The J. Acoust. Soc. Am.* **120**, 343–359 (2006).

21. Majdak, P., Goupell, M. J. & Laback, B. 3-d localization of virtual sound sources: effects of visual environment, pointing method, and training. *Attention, perception, & psychophysics* **72**, 454–469 (2010).

22. Mendonça, C. *et al.* On the improvement of localization accuracy with non-individualized hrtf-based sounds. *J. Audio Eng. Soc.* **60**, 821–830 (2012).

23. Parseihian, G. & Katz, B. F. Rapid head-related transfer function adaptation using a virtual auditory environment. *The J. Acoust. Soc. Am.* **131**, 2948–2957 (2012).

24. Mendonça, C., Campos, G., Dias, P. & Santos, J. A. Learning auditory space: Generalization and long-term effects. *PloS one* **8**, e77900 (2013).

25. Koepp, M. J. *et al.* Evidence for striatal dopamine release during a video game. *Nature* **393**, 266 (1998).

26. Harley, C. W. Norepinephrine and dopamine as learning signals. *Neural plasticity* **11**, 191–204 (2004).

27. Riesenhuber, M. An action video game modifies visual processing. *TRENDS Neurosci.* **27**, 72–74 (2004).

28. Green, C. S. & Bavelier, D. Action video game modifies visual selective attention. *Nature* **423**, 534 (2003).

29. Green, C. S. & Bavelier, D. Action-video-game experience alters the spatial resolution of vision. *Psychol. science* **18**, 88–94 (2007).

30. Li, R., Polat, U., Makous, W. & Bavelier, D. Enhancing the contrast sensitivity function through action video game training. *Nat. neuroscience* **12**, 549 (2009).

31. Honda, A. *et al.* Transfer effects on sound localization performances from playing a virtual three-dimensional auditory game. *Appl. Acoust.* **68**, 885–896 (2007).

32. Lim, S.-j. & Holt, L. L. Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cogn. science* **35**, 1390–1405 (2011).

33. Whitton, J. P., Hancock, K. E., Shannon, J. M. & Polley, D. B. Audiomotor Perceptual Training Enhances Speech Intelligibility in Background Noise. *Curr. Biol.* **0**, DOI: 10.1016/j.cub.2017.09.014 (2017).

34. Zhang, Y.-X., Tang, D.-L., Moore, D. R. & Amitay, S. Supramodal enhancement of auditory perceptual and cognitive learning by video game playing. *Front. psychology* **8**, 1086 (2017).

35. Shinn-Cunningham, B. G., Durlach, N. I. & Held, R. M. Adapting to supernormal auditory localization cues. i. bias and resolution. *The J. Acoust. Soc. Am.* **103**, 3656–3666 (1998).

36. Kumpik, D. P., Kacelnik, O. & King, A. J. Adaptive reweighting of auditory localization cues in response to chronic unilateral earplugging in humans. *J. Neurosci.* **30**, 4883–4894 (2010).

37. Stitt, P., Picinali, L. & Katz, B. F. Auditory accommodation to poorly matched non-individual spectral localization cues through active learning. *Sci. reports* **9**, 1063 (2019).

38. Irving, S. & Moore, D. R. Training sound localization in normal hearing listeners with and without a unilateral ear plug. *Hear. research* **280**, 100–108 (2011).

39. Majdak, P., Walder, T. & Laback, B. Effect of long-term training on sound localization performance with spectrally warped and band-limited head-related transfer functions. *The J. Acoust. Soc. Am.* **134**, 2148–2159 (2013).

40. Amitay, S., Irwin, A. & Moore, D. R. Discrimination learning induced by training with identical stimuli. *Nat. neuroscience* **9**, 1446 (2006).

41. Ortiz, J. A. & Wright, B. A. Contributions of procedure and stimulus learning to early, rapid perceptual improvements. *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 188 (2009).

42. Hawkey, D. J., Amitay, S. & Moore, D. R. Early and rapid perceptual learning. *Nat. neuroscience* **7**, 1055 (2004).

43. Molloy, K., Moore, D. R., Sohoglu, E. & Amitay, S. Less is more: latent learning is maximized by shorter training sessions in auditory perceptual learning. *PloS one* **7**, e36929 (2012).

44. So, R. *et al.* Toward orthogonal non-individualised head-related transfer functions for forward and backward directional sound: cluster analysis and an experimental study. *Ergonomics* **53**, 767–781 (2010).

45. Jones, P. R., Moore, D. R., Amitay, S. & Shub, D. E. Reduction of internal noise in auditory perceptual learning. *The J. Acoust. Soc. Am.* **133**, 970–981 (2013).

46. Aspeslagh, S., Clark, F., Akeroyd, M. A. & Brimijoin, W. Measuring rapid adaptation to complex acoustic environments in normal and hearing-impaired listeners. *The J. Acoust. Soc. Am.* **137**, 2229–2229 (2015).

47. Cuevas-Rodriguez, M. *et al.* An open-source audio renderer for 3d audio with hearing loss and hearing aid simulations. In *Audio Engineering Society Convention 142* (Audio Engineering Society, 2017).

48. Begault, D. R., Wenzel, E. M. & Anderson, M. R. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc.* **49**, 904–916 (2001).

49. Ircam listen hrtf database. http://http://recherche.ircam.fr/equipes/salles/listen/. Accessed: 2019-01-17.

50. Leong, P. & Carlile, S. Methods for spherical data analysis and visualization. *J. neuroscience methods* **80**, 191–200 (1998).

51. Best, V. *et al.* A meta-analysis of localization errors made in the anechoic free field. In *Principles and applications of spatial hearing*, 14–23 (World Scientific, 2011).

52. Blauert, J. *Spatial hearing: the psychophysics of human sound localization* (MIT press, 1997).

53. Katz, B., Rio, E. & Picinali, L. Limsi spatialization engine. *Inter Depos. Digit. Number: F* **1** (2010).

54. Shelton, B. & Searle, C. The influence of vision on the absolute identification of sound-source position. *Percept. & Psychophys.* **28**, 589–596 (1980).

55. Eastgate, R., Picinali, L., Patel, H. & D'Cruz, M. 3d games for tuning and learning about hearing aids. *The Hear. J.* **69**, 30–32 (2016).

56. Steadman, M. A., Kim, C., Lestang, J.-H., Goodman, D. F. M. & Picinali, L. Short-term effects of sound localization training in virtual reality [dataset], DOI: 10.5281/zenodo.2594832 (2019).